

## Paleodemography: Expectancy and False Hope

J.P. BOCQUET-APPEL AND C. MASSET

*Laboratoire d'Anthropologie biologique, Musée de l'Homme, F-75116  
Paris (J.P.B.-A.), CNRS, URA 275 "Méthodes d'Etude des Sépultures,"  
Laboratoire d'Anthropologie, F-33405 Talence (C.M.), France*

**KEY WORDS** Iterative proportional fitting, Age distribution, Demographic estimators

**ABSTRACT** From parent populations ( $N = 50,000$ ) stochastically generated, representing different levels of correlation ( $r$ ) between the age at death and a hypothetical biological indicator ( $r = 0.8-0.98$ ), reference samples and target demographic samples are randomly drawn. Two iterative techniques, *proportional fitting procedure* and *Bayesian*, are used to estimate from the reference samples the age distribution of the targets. Due to the random fluctuations of the pattern of aging, both in the reference and target samples, these techniques converge only in expectation toward the true value of a distribution, but not in practice for any particular realization. Nevertheless, these techniques allow the estimation of the average of an age distribution, even if its shape is unknown. Under the hypothesis that the target sample is drawn from a stationary population, this average represents the life expectancy at 20 years (plus 20 years). Using this mean age at death for the adults and the juvenility index at death ( $D5-14/D20-\omega$ ), a new set of paleodemographic estimators were derived from 40 archaic life tables. For a hypothesized stable population, they give the life expectancy at birth and at 20 years, and the probability of death at 1 and 5 years. © 1996 Wiley-Liss, Inc.

It is now 15 years since we argued that the age distribution of adult skeletons produced by the paleodemographic studies are gravely biased. Our findings suggested that they essentially mirror the age distribution of the collection of the reference skeletons (Masset, 1973; Bocquet-Appel & Masset, 1982; Bocquet-Appel & Masset, 1985; Bocquet-Appel, 1986; Masset, 1982). As a result of our analysis, we first called for a conceptual change: the age distribution of a cemetery must be estimated globally, via a probabilistic approach, and not as a simple sum of the individually estimated age, as is usually done by the traditional method based on forensic medicine. In order to attempt to solve the serious issues with which the paleodemography was confronted, we suggested two approaches: one, from the anthropological field, to give an unbiased estimate for the age distribution of adult skeletons; the other, from demography, to furnish the main demographical param-

eters of a population, despite the absence of a reliable method for estimating the age distribution for the adult skeletons.

To neutralize the influence of the a priori probability of age estimates, directly determined by the anthropological collection's age distribution, we strongly advocated the use of a uniform reference sample for the ages. In the last instance, the number of estimable age classes is conditioned by the information quantity on the age carried globally by the biological indicator(s) and summarized by the correlation coefficient between those age indicator(s) and the age at death. Yet to obtain the main demographical parameters from a cemetery sample, such as life expectancy at birth, and the probability of death of children at 1 and 5 years, we suggested the use of estimators based on 40 reference

---

Received for publication June 10, 1994; accepted July 24, 1995.

tables of mortality, gathered worldwide, and exhibiting a wide mortality experience (Bocquet and Masset, 1977).

Our anthropological conclusion first met strong reaction and led to a great deal of controversy which is still not yet resolved (Van Gerven and Armelagos, 1983; Bocquet-Appel and Masset, 1985; Buikstra and Konigsberg, 1985; Jackes, 1985; Bocquet-Appel, 1986; Piontek and Weber, 1990; Wood et al., 1992; Jackes, 1993); but it stimulated much research and triggered reexamination of the orthodoxy based on forensic medicine, which at one time appeared very firmly established (Bocquet, 1977; Bocquet et al., 1978; Masset, 1982; Jackes, 1985; for a review see Jackes, 1992). Other researchers have reached analogous conclusions (Jackes, 1985; Langenscheidt, 1985; Molleson and Cohen, 1990; Konigsberg and Frankenberg, 1992; Aiello and Molleson, 1993). As far as our demographic suggestion is concerned—the use of estimators—there was little reaction (Jackes, 1992, 1993), probably because of the undue confidence in the adult age distribution produced by the anthropologists.

In their paper, Konigsberg and Frankenberg (1992) described our recommendation to use a uniform reference sample as problematic since “a reference sample with a uniform age distribution (...) requires discarding data, which certainly cannot be an efficient way to proceed” (p. 239). They suggest using an iterative Bayesian procedure (IB) published by Kimura and Chikuni (1987). The results obtained by Konigsberg and Frankenberg from simulated data are fairly good. Obviously, we never suggested discarding data, but rather to base the estimation technique on an a priori uniform distribution (Bocquet, 1977; Bocquet-Appel, 1986), because what we had in mind at that time was to employ a method of estimation that was technically quite close to the iterative Bayesian procedure advocated by Konigsberg and Frankenberg. Indeed, in 1977 we experimented with such a method, known as *iterative proportional fitting procedure* (IPFP). Except in summary (Masset, 1982), we did not publish it in full, because despite appearances, as we shall see, neither the IPFP nor the IB method give the solution. The critical issue is not the mathematical

technique of estimation, but the random sampling for the pattern of aging. Generally, even if the estimates converge in expectancy, based on a single reference sample, it is not possible to use them to estimate the age distribution of a target sample for a particular outcome. But Konigsberg and Frankenberg's paper triggered reexamination of the iterative technique.

In this article, we test again the approach proposed by Konigsberg and Frankenberg (1992). We will focus our paper on the range of the error variance in the adult age distribution viewed as the variance of histograms representing the age distributions generated by the two iterative estimation techniques (IPFP and IB). Further, we will use these techniques to estimate accurately not the shape of an adult age distribution, but its average, which is in fact in a sample of a stationary population, the life expectancy at 20 years (plus 20 years). As a natural extension of these iterative techniques, new groups of paleodemographic estimators will be recomputed.

### A COMPARISON OF TWO ITERATIVE PROCEDURES OF ESTIMATION: PROPORTIONAL FITTING AND BAYESIAN

In 1977 we rediscovered an iterative fitting technique already known in the statistical literature and called iterative proportional fitting procedure (IPFP). This technique was originally conceived by Deming and Stephan (1940), for estimating the cell frequencies of a joint distribution in a contingency table when the marginal ones are known. This technique paved the way for several developments in the analysis of contingency tables and model fitting (see, among others, Fienberg, 1970; Haberman, 1974; Meyer, 1982). The issue was the following: knowing the joint distribution of a biological indicator and the age at death in a sample, but only one of the two derived marginal distributions, could we use the IPFP technique to estimate the marginal distribution which was missing? In this paper, we will call the *reference sample*, the anthropological collection for which the joint distribution,  $f(a,i)$ , is known; and the *target sample*, as do Konigsberg and Frankenberg (1992),

the sample for which only the biological indicator distribution,  $f(i)$ , is known and the age class distribution,  $f(a)$ , has to be estimated, where  $a$  and  $i$  represent, respectively, the age and the biological indicator.

In a *reference sample*, let  $n_{ia} = n_{ai}$  be the frequencies of individuals in the  $a$ -th age class having the  $i$ -th stage of an indicator distributed in a table at  $r$  rows and  $c$  columns ( $i = 1, \dots, r$ ;  $a = 1, \dots, c$ ). In this table ( $r \times c$ ), one has two marginal sums, in column

$$n_a = \sum_i n_{ia}$$

and in row

$$n_i = \sum_a n_{ia}$$

with  $n_a$  and  $n_i$  representing respectively the frequency of individuals in the  $a$ -th age class and the  $i$ -th stage of an indicator. If one weighs the frequency,  $n_{ia}$ , by the column total (the age), one will have the stage proportion respective to the age class:  $\mathbf{F} = (f_{ia}) = n_{ia}/n_a$ , where  $\mathbf{F}$  represents the table  $r \times c$ , of element  $f_{ia}$ , where  $f_{ia}$  is the observed proportion of the  $i$ -th stage for the  $a$ -th age class. The column total is

$$f_{\cdot a} = \sum_i f_{ia} = 1$$

for all age-classes.

The array  $\mathbf{F}$  can be read in two ways: on the one hand, it represents what is called a conditional distribution (of the stages given the age classes) in statistics; and on the other hand it represents the distribution of a reference sample in which all age classes have the same frequency, for which the value is  $f_{\cdot a} = 1$ . This is where the idea to construct a uniform reference sample comes from, which was unfortunately interpreted as discarding data by Konigsberg and Frankenberg (1992), which is certainly not the case! Whatever the shape of the distribution of the age classes for the reference sample (gaussian, assymmetric, bimodal, irregular, etc.) a simple

technical trick—dividing the observed number  $\{n_{ia}\}$  by the corresponding column total,  $n_{ia}/n_a$ —renders that distribution uniform. A useful quantity in the IPFP algorithm we used is also

$$f_i = \sum_a f_{ia}$$

where  $f_i$  is a marginal quantity of the  $i$ -th stage in the  $\mathbf{F}$  array.

In a *target sample*, let  $m_i$ ,  $m_a$ , and  $m$  be, respectively, the *known frequencies* of individuals in the  $i$ -th stage of an indicator, the estimated frequencies in the  $a$ -th age class, and the total. Ignoring the true distribution of the sample in the age classes, it is legitimate as a first approximation to distribute the sample uniformly, i.e., to give the estimated frequencies, the value  $m_a^0 = m/c$ , where  $^0$  means before all iteration. Our algorithm involves four steps:

$$1. \quad f_{ai}^{(t)} = f_{ia}^{(t-1)} / f_i^{(t-1)}$$

where  $f_{ai}^{(t)}$  represents a standardized distribution by its marginal quantity of the  $a$ -th age class for the  $i$ -th stage of the indicator;

$$2. \quad m_a^{(t)} = \sum_i f_{ai}^{(t)} m_i$$

where  $m_a^{(t)}$  represents the estimated frequency in the  $a$ -th age class, given  $m_i$  the known marginal distribution of the individuals in the  $i$ -th stage for the target;

$$3. \quad f_{ia}^{(t)} = f_{ia}^{(t-1)} m_a^{(t)} / m_a^{(t-1)}$$

where  $f_{ia}^{(t)}$  represents a new iterated-value of the preceding  $f_{ia}^{(t-1)}$ ,

$$4. \quad f_i^{(t)} = \sum_a f_{ia}^{(t)}$$

where  $f_i^{(t)}$  represents the new value for the  $i$ -th stage in the corresponding marginal distribution.

The procedure stops when the  $m_a$ -values remain practically unchanged between two

successive iterations or when the sum of squares

$$S^{(t)} = \sum_{a=1}^c (m_a^{(t)} - m_a^{(t-1)})^2$$

does not decrease any more. A version of an old Fortran program (ca. 1977) for that algorithm can be found in Masset (1982).

### Simulated data: Parent populations and target samples

Four bivariate distributions, approximately uniform, representing the stages of an indicator (seven stages) and the age at death (from 20 to 90 years) in a parent population, have been randomly generated, following the procedure described by Kennedy and Gentle (1980). In our simulation, instead of introducing univariate normal distributions, to get the corresponding multivariate ones, via a Cholesky decomposition of a patterned covariance matrix, we have used univariate uniform distributions. The resulting distributions are only approximately uniform, but this is enough for the purpose of this paper. A point is in order here. If we used a uniform distribution in these simulations, for the parent populations, this is *not* for the same reason as the main issue of this article, namely, to avoid the influence of the reference sample. As was shown above, the simple construction of a  $\mathbf{F}$  matrix renders *any* crude age distribution uniform, whatever its original shape. Here, we used a uniform distribution in order to avoid any contamination from a minor issue: a heterogeneity between age classes for the precision of the estimate of the  $\mathbf{F} = \{f_{ia}\}$  matrix. That precision being proportional to  $[n_a f_{ia} (1 - f_{ia})]^{1/2}$ , it can vary from age class to age class as their sample size,  $\{n_a\}$ , so generating an heterogeneity of precision *within* the  $\mathbf{F}$  matrix (between its column elements). This is to escape a possible influence of that heterogeneity on the simulations that we tried to eliminate it at first, by using approximately identical  $n_a$ , i.e., a uniform distribution for the age classes. As we will see later, in practice this heterogeneity does not have a significant influence on the result if the sample sizes for the age classes are not too

small. The parent populations have been generated for four correlation levels between the age and the biological indicator,  $r(a, i)$  of 0.8, 0.9, 0.95, and 0.98. Each of the four parent populations comprises 50,000 individuals. Table 1 shows an example of their joint distributions in seven stages of a hypothetical age indicator and seven age classes.

Three target samples, representing various models of age distributions which could possibly be found in cemeteries, have been formed by random sampling with replacement in each of the four parent populations: unimodal (old regime demography), bimodal (the usual mode for elderlies plus a complementary mode for violent death and maternal overmortality), and triangular (a war death cemetery) (Fig. 1). In order to get stable estimates for the statistical parameters sought, the simulations have been repeated five times. In all, we have three target samples  $\times$  five duplications  $\times$  four correlation levels: 60 different target samples.

### Expectancy and variance of the estimates on the target samples

Konigsberg and Frankenberg were interested only in the expectancy of the estimates for age class frequencies. But, in practice, for a given target, the anthropologist has one, and *only one* reference sample at his disposal. He can only obtain one estimate, which practically, as we shall see below, never corresponds to the expectancy. The issue is in both techniques: besides the expectancy of an estimate, what is its variance of error?

To produce a variance of error, a reference sample, uniformly distributed in the age classes, is generated by randomly drawing, with replacement,  $N_R$  individuals from a parent population. From that reference sample (its  $\mathbf{F}$  table) and the marginal distribution in the target samples ( $m_i$ ), an estimate of the age distribution is computed, by the two techniques, IPFP and IB, for the three target samples (old regime, bimodal, triangular). This operation is repeated 500 times. In these simulations of error variance, the reference and the target samples, although random, have both a fixed shape for their distributions, one being constantly uniform, the other "demographic." Later, we will general-

TABLE 1. A simulated joint distribution for a parent population of a hypothetical age indicator and the adult's age at death, cross-classified in 7 stages and 7 age classes ( $r = 0.796$ )

Indicator stage	Age							Total
	20-29	30-39	40-49	50-59	60-69	70-79	80-89	
I	1,519	207	0	0	0	0	0	1,726
II	2,827	2,767	1,466	149	0	0	0	7,209
III	2,629	2,892	2,971	2,765	1,306	119	0	12,682
IV	108	1,246	2,687	2,919	2,957	2,711	1,165	13,793
V	0	0	148	1,389	2,810	2,949	2,687	9,983
VI	0	0	0	0	178	1,573	2,644	4,395
VII	0	0	0	0	0	0	212	212
Total	7,083	7,112	7,272	7,222	7,251	7,352	6,708	50,000

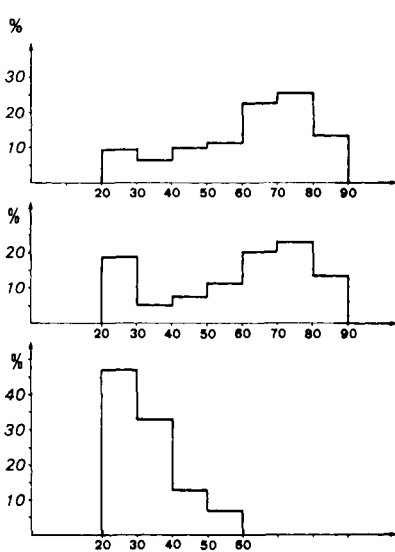


Fig. 1. Three target distributions. From top to bottom: old regime demography, bimodal, triangular.

ize the approach by allowing the shape of the distributions to vary in both samples. There are at least two ways of representing the variance of error: one is strictly numerical—a sum of square deviations from the true frequency-values—; the other by a histogram distribution, each of the histogram in the distribution being the graphical version of the age class estimates for the target samples.

A numerical estimate of the sampling variance,  $V(p_c)$  can be

$$V(p_a) = \frac{1}{Q} \sum_k^Q (p_{ak} - p_a)^2 \quad (a = 1, \dots, c-1)$$

$$V(p_c) = \sum_a^{c-1} V(p_a)$$

where  $V(p_a)$ ,  $p_{ak}$ ,  $p_a$  and  $Q$  represent, respectively, the sampling variance of the frequency estimate for the  $a$ -th age class, the observed estimate at the  $k$ -th simulation ( $k = 1, \dots, Q$ ), the true frequency, and the number of simulations.

An age distribution is currently depicted by a histogram where the frequency of individuals is on the vertical axis, and the age classes on the horizontal one. Independent of the absolute values of the frequencies of the age classes, such a histogram is interesting in itself because one can hypothesize the kind of demographical profile to which the sample belongs by a simple visual inspection or by statistical model fitting such as the competing hazard model (Siler, 1979, 1983; Gage, 1990, 1991). In order to compare a reasonable number of different histograms, we shall focus our attention not in the absolute value of the frequencies in the age classes but in their ranking. In this approach one considers that two histograms are identical if their age classes have the same ranking, independently of the absolute values of the individual frequencies; for instance, the two histograms representing the three frequencies ( $a_1 = 60$ ,  $a_2 = 30$ ,  $a_3 = 10$ ) and ( $a_1 = 80$ ,  $a_2 = 15$ ,  $a_3 = 5$ ) are considered identical because the ranking of their age classes ( $a_1 > a_2 > a_3$ ) is the same.

For a histogram at  $c$  age classes,  $c!$  different hierarchies of classes are possible, excluding the hypothesis of equal frequencies distribution between two or several age classes. In that way the variance of the esti-

mates can be viewed now as a distribution of histograms among the  $c!$  different possible histograms. In this approach one is interested in the realization frequency of the true histogram, among all the histograms generated, these latter representing the variance sampling around the true histogram.

The importance of the error variance depends on the importance of the sampling fluctuations in the reference and target samples simultaneously. These fluctuations are a function of the three following quantities: 1) the correlation between the pattern of aging and the age at death,  $r(i, a)$ ; 2) the size of the reference,  $N_R$ , and the target samples,  $N_T$ ; and 3) the number of the age classes in the target sample (Bocquet, 1978; Bocquet-Appel and Masset, 1982). In the second part of this paper, we will summarize the main causes for the error variance using iterative techniques.

This is the overall quantity of information in the age at death carried by one (or several) indicator(s), measured by  $r(i, a)$ , which is by far the most important and not the number of the indicator itself, as Konigsberg and Frankenberg (1992) believe. A single indicator could carry as much as or even more information than several correlated ones. The greater the information quantity, the lesser the influence of the sampling fluctuation in the pattern of aging for the reference and target samples. For the limiting value of a correlation of  $r(i, a) = 1$ , we will get a totally determined system, whatever the values of  $N_R$ ,  $N_T$  and  $c$  may be. After a sweeping of our 60 simulations, we finally retained a domain of estimates determined by the following realistic parameter-values:

$$r(i, a) = 0.8, 0.9, N_R = 700, \\ N_T = 500, c = 4, 7.$$

We shall compare, for the three target samples the following results, given by the IPFP and IB technique:

i. *The expectation of the estimates.* In both techniques that quantity, given by the grand

average is rather close to the true value of the age classes (Figs. 2, 3).

ii. *The realization frequency of the true histogram for the age class distribution* (Table 2, column 2). For the two demographic target samples (old regime, bimodal), the true histogram has been produced, on the average, in 45% (IB) and 41% (IPFP) of the simulations. If the number of age classes on the histogram is increased, the result gets rapidly worse. With a seven-age-class histogram (not shown here in order to save space), the true histogram is practically never produced (below 1%). For the triangular distribution, the results are fairly good, with a realization rate of about 95%.

iii. *The detection frequency of the modal age class for the true histogram* (Table 2, column 3). For the two demographic target samples, the modal class is produced in 75% (IB) and 68% (IPFP) of all histograms generated, and in 100% for the triangular target distribution.

Whatever the technique, IPFP or IB, it clearly can be seen that the realization frequency of the true histogram is shape-dependent on the target sample distribution. The most easily realized distribution has a very simple shape of monotonically ordered classes with only one mode. The hardest is a bimodal distribution, i.e., one that has two maxima. With seven age classes, instead of four (not shown here), the detection of such a bimodal distribution, where one mode is located on one side while the other is lying somewhere between the two sides, has proved impossible unless the correlation level is as high as  $r(i, a) = 0.98$ . The realization rate is practically zero in both techniques below that correlation level. But the important conclusion is, even if the expectancy of the estimates converges toward the true values of the age-class distribution, the true histogram (with only four classes) is obtained in less than 50% of the cases on average, discarding the triangular distribution. There is less than one chance in two that the observed histogram is the true one. When it is, there are no criteria to prove it.

Bootstrapping the reference sample in order to generate an expectation for the estimates would be a wrong approach, because

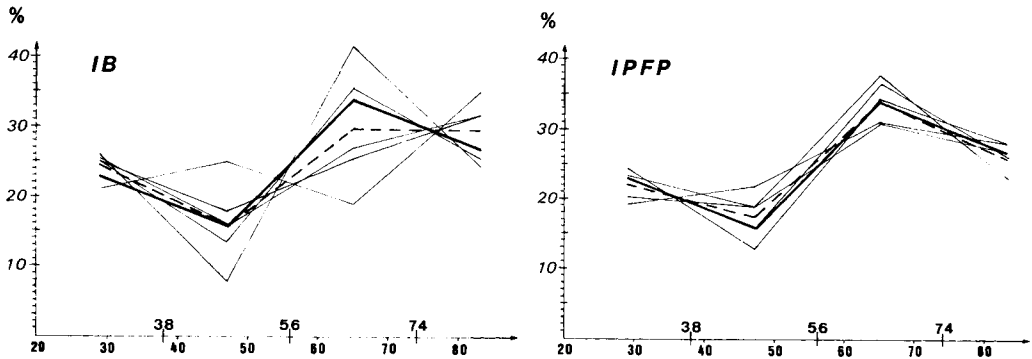


Fig. 2. Age distributions of target and estimated samples, in a bimodal case for the target, with 4 age classes. Expected correlation ( $r$ ) between age indicator and age = 0.8. To improve readability, histograms are replaced by segments of lines connecting the middles of classes. Horizontal axis, ages; vertical axis, age class frequency; IB, iterative bayesian procedure; IPFP, itera-

tive proportional fitting procedure; thick line, target sample; **thin lines**, average of 500 estimates (each one obtained by iterations), **broken line**, grand average. Convergence appears on the average but not in practice for any particular individual estimate. The difference between the IB (left) and IPFP (right) for the figures seems to be due to random fluctuations.

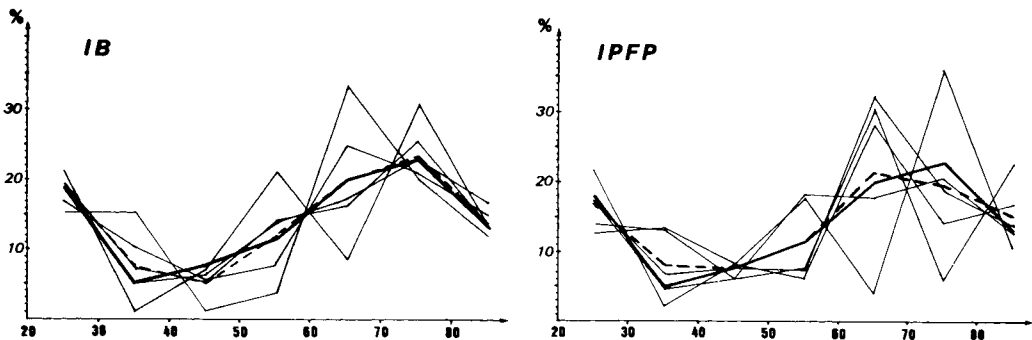


Fig. 3. Age distributions of target and estimated samples, in a bimodal case, using 7 age classes. The expected correlation ( $r$ ) between age indicator and age is 0.9. The results of Figures 2 and 3 are comparable, but the ranking of classes is more often found in the case of Figure 1 (see text for explanation).

the expectation so generated would be in the expectancy domain of the reference sample. In fact, the reference and the target sample are both samplings of the unknown parent population. The target sample by no means derives from the reference sample as did the simulation model of Konigsberg and Frankenberg (1992). Both samples are not parent and child but siblings.

If the expectation computation has raised some hope, unfortunately it should be abandoned, simply because the expectation is not estimable in that context. A unique estimate has practically no chance to depict the reality, due to sampling fluctuations both in ref-

erence and target samples. This is also the view of Kimura and Chikuni (1987), the authors of the IB method, who concluded their paper this way: "The critical assumptions of the . . . method are that the age-length [indicator] data have been adequately sampled, and that the associated  $q_{ij}$  [prob (indicator/age)] estimates are applicable to the population sampled for length [indicator] frequencies. The latter may be a poor assumption since the age-length [indicator] data may be from a different year, time period, or geographic area than the sampled length [indicator] frequencies. . ." (Kimura and Chikuni, 1987, p. 34). In light of our simulations, only

TABLE 2. Simulated error variance, using the IPFP and IB technique, for the estimates of an age class distribution of 4 classes, in three target sample distributions

Correlation age-indicator	Target distribution	Error variance <sup>1</sup>		Percent of true histogram realized		Percent of true detected mode	
		IPFP	IB	IPFP	IB	IPFP	IB
0.80	Old regime	.4582	.4183	30.7	40.	68.6	71.8
	Bimodal	.4300	.4306	30	21	73.9	48.9
	Triangular	.3922	.5688	95.5	84.8	100.	100
0.90	Old regime	.3245	.2315	57.2	57.8	59.7	87.9
	Bimodal	.1716	.2548	45.4	61.2	71.9	90.8
	Triangular	.3229	.2442	92.4	100.	100.	100.

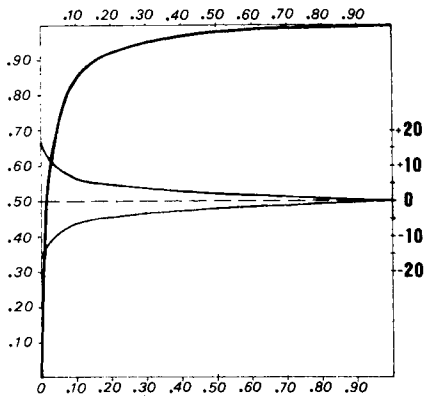
<sup>1</sup> x E-02.

Fig. 4. Relationship between the true and estimated mean age of target samples obtained by simulation (iterative Bayesian method). Horizontal axis, correlation between age indicator and age; vertical axis (left), correlation between the true and estimated mean age of the target samples. The thick curve gives the average correlation for 5 duplicates (500 target samples drawn by duplicate). Vertical axis (right), standard error in years given the correlation level between age indicator and age. The standard error is represented by the thin symmetrical curves.

those who read between the lines will understand: no hope with the expectation.

### A NEW APPROACH: THE ESTIMATE OF THE AVERAGE FOR ADULT AGE DISTRIBUTION AT DEATH

Although it is not possible to estimate reasonably the shape of an age distribution by the iterative techniques described above, one can nevertheless obtain a fairly good estimation for the direction of the overall trend of an undefined distribution: its average.

Figure 4 shows the correlation between the true and estimated average age, pro-

duced by both iterative techniques, for random target and uniform reference distributions and the standard error for the estimate, for different correlation levels between the age and the biological indicator. Here the two iterative techniques give the same result. Those results were obtained using the following simulation device: a reference ( $N_R = 700$ ) and a target ( $N_T = 500$ ) sample are generated stochastically, with replacement, from a parent population. The shape of the target sample is random, including the case where one or several age classes among  $c - 1$  might have a sample size of zero. The estimate for the average of the random distribution is obtained, using one or the other technique, IPFP or IB, along with the **F** matrix formed from the reference sample (see above). Five hundred pairs of reference and target samples are drawn. The correlation between averages is computed together with its standard error. For each of the four correlation levels between the age and the indicator, the procedure is repeated five times. With a biological indicator, represented by a correlation with the age at death as low as 0.60 and for  $N_T = 500$ , the number of adult skeletons in the cemetery, the standard error for the estimation of the average age of the unknown distribution is 2.09 years, and the correlation coefficient between the true and estimated average is 0.98. With  $N_T = 100$ , the simulated standard error is 3.05 years (Bocquet-Appel, 1994). For the same level of correlation age-indicator, the number of columns for the **F** matrix, i.e. the number of age-classes from 7 to 4, does not seem to influence the result.

These simulations were done with refer-



TABLE 3. Variation of the correlation between true and estimated age by the iterative technique, in three shapes for the age-class distribution: random, constrained, uniform<sup>1</sup>

Pattern of aging $r(i,a)$	Shape for the age-class distribution			
	Random		Constrained	Uniform
	IB	IPFP		
0.60	0.921	0.949	0.981	0.983
0.90	0.972	0.972	0.997	0.997

<sup>1</sup> $N_R = 700$ ,  $N_T = 500$ . See text for explanation. IB = iterative bayesian, IPFP = iterative proportional fitting procedure.

ence and target samples of equal and relatively large fixed size. But does the observed correlation between the true and estimated average age remain stable if the shape of the age distribution for the reference sample varies? A variation between the sample sizes of the age classes in a reference induces a heterogeneity of precision in the **F**-matrix, i.e., between its column elements, which, in turn, may bias the estimates. Table 3 shows the observed correlations between true and estimated age, for two patterns of aging in the parent population, and in three types of age distributions in the reference sample: i) uniform, ii) random, and iii) constrained. In the random distribution, one or several of the  $c-1$  age classes could have a size equal to zero. In the constrained distribution, the minimum frequency of individuals in an age class has been arbitrarily set to  $1/2 c$ —i.e., there are at least  $N_R/2\alpha$  individuals in an age class. This eliminates the unrealistic situation for a “reference” sample of having no individual in one or even several age classes, while preserving a reasonably large heterogeneity among the age-class frequencies, like those observed in the anthropological collections of Coimbra (Bocquet-Appel and Xavier de Morais, 1987) and Spitafeld (Molleson and Cox, 1993). The correlations given in Table 3 were obtained using the same simulation device described above. For the two references, uniform and constrained, the correlation between true and estimated average age remains the same. But it is weaker for the random age distribution. Here the two iterative techniques do not give exactly the same result, but they are not very different (for  $r(i,a) = 0.60$ : 0.921 with IB and 0.949 with IPFP). In short, Table 3 shows that if the simulated constrained sample is repre-

TABLE 4. Standard error for the average age produced by the iterative techniques as a function of the variation of size in the reference and in the target samples<sup>1</sup>

Target sample size	Reference sample size			
	100	250	500	1,000
50	5.066	4.392	4.122	3.873
	2.009	1.688	1.516	1.400
100	4.431	3.658	3.289	3.038
	1.792	1.407	1.190	1.143
250	3.951	2.942	2.470	2.142
	1.555	1.130	0.960	0.852
500	3.680	2.613	2.093	1.764
	1.440	1.014	0.827	0.727
1,000	3.638	2.428	1.877	1.497
	1.442	0.954	0.791	0.670

<sup>1</sup>The age-class distribution is random for the target and constrained for the reference. For the first and second line,  $r(i,a)$  has a value of, respectively, 0.60 and 0.90.

sentative of major anthropological collections of reasonable size and sampling, like Coimbra and Spitafeld, the heterogeneity of distribution among their age classes should not noticeably affect the estimates for the average age of an unknown age distribution produced by the iterative techniques.

In a paleodemographic situation, to estimate the standard error for the average age, the variation in sample sizes have to be taken into account, both in the reference and in the target. Table 4 shows the standard error when both sample sizes vary simultaneously (from 100 to 1,000 for  $N_R$ ; and from 50 to 1,000 for  $N_T$ ). The procedure used to generate the standard error is, again, the same as that described above. With a pattern of aging for the biological indicator characterized by  $r(i,a) = 0.60$ ,  $N_T = 100$  and  $N_R = 500$ , the simulated standard error is 3.29 years; with a better indicator ( $r(i,a) = 0.90$ ) the error is 1.19 years. Other values, for different patterns of aging, may be found in Bocquet-Appel (1994).

Finally, the error for the mean age may be summarized as being due to

$$\begin{aligned}
 \text{error for the mean age} &= \text{error for the pattern of aging} + \text{error for the reference sample} + \text{error for the cemetery sample} \\
 &\quad + \text{unexplained residual} \\
 &= \text{error in the lack of information redundancy age-indicator} + \text{error in the sampling for the F-matrix} + \text{error in the sampling for the target } (m_i) \\
 &\quad + \text{technical}
 \end{aligned}$$

The error due to the pattern of aging takes into account the lack of information redundancy between the age and the biological indicator(s), which is characterized by  $1 - r(i, a)^2$ , the simple or multiple correlation coefficient. The error due to the reference takes into account the sampling for the **F**-matrix, relative to the unknown **F**-matrix for the parent population. It includes the limited size of the sample along with the unequal distribution for the age-classes in the reference. The error due to the cemetery sample takes into account the sampling for the  $(m_i)$  distribution, the stages for the biological indicator, relative to the unknown "demographic" distribution for the parent population, due again to the limited size of the target. Finally, we called the "technical" residual that small difference which can be observed when one estimates the average age of the target by its own **F**-matrix, for which there is no clear explanation.

Is there, at least, a condition of convergence towards the solution for these iterative techniques? The simulations have shown that the pattern of the **F**-matrix should represent a continued increase with the age of the value taken by the biological indicator. Let  $X_a$  be the conditional mean for the stage of an indicator in the age class  $a$ ,  $X_a = \sum f_{ia} x_i$ , where  $x_i$  represents the value taken by the  $i$ -th stage of the indicator (generally an integer) and  $(X_1, X_2, \dots, X_c)$  the set of conditional means derived from the **F**-matrix. The simulations have shown that the computations will converge if the conditional means displayed increasing values:  $X_1 < X_2 < \dots < X_c$ . The increase between two successive means,  $\Delta x$ , do not need to be regular. This fact is important. Simply, this means that, if an overall correlation age-indicator is a prerequisite condition for the iterative technique to converge, the within-age-class correlation does not need to be homogeneous (or "uniform") for the whole life span. This is the meaning of the variations for the  $\Delta x$ 's. Put otherwise, it is not necessary that the conditional means be aligned on the same correlation slope. In practice, a proper handling of the data on the rows or/and on the columns, i.e., lumping together neighbouring age classes or stages, makes it easy to meet this condition.

## PALEODEMOGRAPHIC ESTIMATORS

In our 1982 paper, we suggested another approach, i.e., by employing "estimators" enabling their user to appraise some demographic parameters from the ratio between the number of dead from 5 to 14 and the number of dead after 20 ( $D_{5-14}/D_{20-\omega}$ ). This juvenile/adult ratio is a juvenility index at death (Bocquet and Masset 1977; Bocquet 1979). To establish these estimators we did not start from model life tables of the demographic literature, because no table exists corresponding to our desiderata. Among those constructed solely from controlled data (Ledermann, 1969; Coale and Demeny, 1983), none allows as input the ratio  $D_{5-14}/D_{20-\omega}$ . Furthermore, to construct their "west model," Coale and Demeny deliberately dropped all historic tables prior to 1870 (Coale and Demeny 1983:12), a date too recent for the study of archaic mortality. So we had to start anew from raw data. Since we were interested in a mortality pattern of a pre-jennerian type (that is, prior to the progress in vaccinations and public health associated with Jenner), we based our calculations on 40 tables from demographically archaic populations, embodying a wide spatial experience of mortality (Bocquet and Masset, 1977).

Establishing estimators supposes a minimum of regularity in human population demography, even in remote antiquity. We find here once more the "uniformitarian hypothesis" without which paleodemography cannot be conceived. In the case of these estimators, a requirement to be satisfied immediately concerns maximal longevity, a parameter which, some two or three centuries ago, reached values well above those with which it is credited in most of paleodemographic literature. In other words was there, in the 15th century A.D., a silent demographic revolution between the last cemeteries where longevity would not have reached 70, if we trust most paleodemographers, and the earliest parish registers, where it almost always exceeded 90?

To answer this question we had to call upon Cutler's work (1975), which completes that of Sacher (1975). The latter has shown that in all Placentalia there exists a strong

TABLE 5. Age estimations of sexual maturity and longevity, as a function of brain and body weights, after Cutler (1975)

Primates (males)	Sexual maturity		Potential longevity	
	Computed	Observed	Computed	Observed
<i>Presbytis</i> (Langur)	3-4	4-5	22	24
<i>Papio</i> (Baboon)	5-6	6-7	32	35
<i>Hylobates</i> (Gibbon)	7-8	6-7	32	30
<i>Pan troglodytes</i> (chimpanzee)	8-9	8-9	46	45
<i>H. sapiens</i> (XVII-XVIII century)	17-18	17-18	92	95

statistical relationship between the weight of the brain controlled by body weight on the one hand, and sexual maturity and longevity on the other hand. Here "longevity" is a potentiality, an estimation of the normal duration of life as obtained from Gompertz's formula (a formula assuming that, starting from the end of childhood, mortality grows exponentially). The regression propounded by Cutler is well verified in all extant Primates including *Homo* (Table 5).

A drawback of these estimators is that they require that a hypothesis on the rate of increase be fixed for the skeletal population, a hypothesis that cannot be deduced from the distribution of skeletons by itself, but which must be based on archaeological evidence. Also, their precision is, as usual, a function of the sample size of the cemetery (Masset and Parzysz, 1985). But, on the whole, their results appear satisfactory when compared to controlled data: sample skeletons from a cemetery and its parish register (Guy and Wabont, 1994).

This situation is improved with the new approach to estimating with some accuracy the mean age at death of a buried set of skeletons. Remember that whatever the parameter-value computable from an age distribution at death, like the index of juvenility or the mean age of a skeleton distribution (Johansson and Horowitz, 1986), it will take a demographical meaning only if one is able to fix the rate of increase corresponding to the unknown living population. Under the hypothesis of a stationary population (rate of increase = 0) the mean age of a distribution of death from 20 and older is nothing else than the life expectancy at age 20 plus 20 years. Therefore new groups of estimators were recomputed, which may be considered as a natural extension of the iterative technique described above. We generated 840

stable populations from our sample of 40 archaic life-tables, for a rate of increase in the range of  $\pm 2.5\%$ , at a step of  $0.25\%$ , following the procedure described by Bourgeois-Pichat (1994). The three variables, one plus the rate of increase ( $1 + \rho$ ), the inverse of the square root for the index of juvenility ( $IJ$ )<sup>-1,2</sup>, and the mean age at death for adults (20 years and more:  $a_{20-w}$ ) were raised up to the power 5 and their products were computed. From these were selected, by stepwise multiple regression (forward selection), the best set of predictors for each of the three following groups of estimators:

i. from the mean age at death for adults ( $a_{20-w}$ ): the life expectancy at 20 years ( $e_{20}$ ). The multiple correlation between  $e_{20}$  of the tables and that  $\hat{e}_{20}$  estimated by the regression is  $R(e_{20}\hat{e}_{20}) = 0.997$ ;

ii. from the index of juvenility ( $IJ$ ): the life expectancy at birth ( $e_0$ ;  $R(e_0\hat{e}_0) = 0.937$ ), the probability of death at 1 year ( ${}_1q_0$ ;  $R({}_1q_{01q_0}) = 0.850$ ) and at 5 years ( ${}_5q_0$ ;  $R({}_5q_{05q_0}) = 0.815$ );

iii. from both, the mean age at death for adults ( $a_{20-w}$ ) and the index of juvenility ( $IJ$ ): the life expectancy at birth ( $e_0$ ;  $R(e_0\hat{e}_0) = 0.952$ ), the probability of death at 1 year ( ${}_1q_{01q_0} = 0.857$ ) and at 5 years ( ${}_5q_{05q_0} = 0.889$ ). Relative to group ii, there is a small improvement here.

Those three groups of estimators are shown in Table 4. For the first group of estimators, there is a very small bias (0.5–0.25 years) at the bounds of its practical use ( $15-20 \text{ years} \leq e_0 \leq 30-35 \text{ years}$ ). It is probably due to a non-uniformity for the shape of the polynomial function for the whole range of  $\rho$  ( $\pm 2.5\%$ ). By taking a number of degrees of freedom equal to the 40 life tables (and not to the 840, because 800 of them

TABLE 6. Paleodemographic estimators<sup>1</sup>

Variables in input: $x_1 = (1 + \rho)$ , $x_2 = a_{20-\infty}$ , $x_3 = (IJ)^{-1/2}$	
i) With the mean age at death at 20 years and more: $e_{20} = -87.3733 + 64.8761x_1^3 + 0.105E-01x_1^2x_2 \pm 0.2146$	$R^2 = 0.995$
ii) With the index of juvenility: $e_0 = -0.474.7660 + 456.4114x_1 - 1.6469x_3^2 + 22.8708x_1^5x_3 \pm 1.521$ ${}_1q_0 = 3.2667 - 2.7286x_1 + 0.101E-01x_3^2 - 0.139x_1^5x_3 \pm 0.0151$ ${}_5q_0 = 7.1968 - 6.1388x_1 - 0.3449x_1^5x_3 + 0.314E-01x_1^5x_3^2 \pm 0.0368$	$R^2 = 0.878$ $R^2 = 0.724$ $R^2 = 0.665$
iii) With the mean age at death at 20 years and more and the index of juvenility: $e_0 = -232.8717 + 168.0674x_1^3 + 1.1678x_1^2x_2 - 0.347E-02x_3x_2^2 + 22.3675x_1^2x_3 \pm 1.3215$ ${}_1q_0 = 0.9093 + 0.162E-05x_2^3 + 0.437E-02x_3^2 - 0.120E-01x_1^5x_2 - 0.116x_1^5x_3 \pm 0.0142$ ${}_5q_0 = 2.7086 - 1.604x_1^3 + 0.557E-04x_3x_2^2 - 0.584E-06x_1^5x_3^3 - 0.406x_1^5x_3 \pm 0.029$	$R^2 = 0.908$ $R^2 = 0.734$ $R^2 = 0.790$

<sup>1</sup> Demographic parameters:  $e_0$ , life expectancy at birth;  $e_{20}$ , life expectancy at 20 years;  ${}_1q_0$ , probability of death at one year;  ${}_5q_0$ , probability of death at 5 years;  $\rho$ , rate of increase. Archeological parameters:  $a_{20-\infty}$ , mean age at death at 20 years and more; IJ, index of juvenility (D 5-14/D 20- $\infty$ )

were generated from the original 40) and to the number of variables used in each multiple regression (both minus two), the rejection levels for the null hypothesis ( $R = 0$ ) are all  $P < 0.001$ . Compared to those we published in 1977, these new estimators are easier to use because they directly incorporate the hypothesis for the rate of increase ( $1 + \rho$ ). Except for the estimator in group i, which is a natural extension of the iterative technique and did not exist before, the precision for the three demographic parameters is identical to that we published in 1977 in group ii but better in group iii.

Finally, with good age indicators we should be able to trust also to a certain extent the modal age class at death obtained by the iterative technique. For instance, with a correlation of  $r = 0.90$  between the age and the indicators, the true modal age class (4 classes; Table 2) is produced in about 70% of the estimates.

### CONCLUSIONS

The iterative techniques are emancipated from the influences of the a priori probability of the reference samples; but, due to the random fluctuations of the pattern of aging, both in the reference and target samples, those techniques only converge in expectancy toward the true values of a distribution. They do not permit reliable assessment of the shape of a distribution.

Nevertheless, those techniques allow accurate estimate of the average of an age distribution, whatever its original shape, even for sample size in a cemetery as small as 100 and a correlation level as low as 0.60 between the age and the biological indicator.

Under the hypothesis that the target sample is drawn from a stationary population, this average represents the life expectancy at 20 years (plus 20 years). We think this finding, combined with the new set of demographic estimators, opens a new door to studies of the skeletal populations. It allows demographic study when the age distribution is unknown.

### ACKNOWLEDGMENTS

We are indebted to Jean Noël Bacro, Richard Paine, and two anonymous referees for valuable comments on an early draft. The manuscript was prepared by Marie-France Leroy, Monique Tersis, and Renée Angelina Hublin.

### LITERATURE CITED

- Aiello LC, Molleson T (1993) Are microscopic ageing techniques more accurate than macroscopic ageing Techniques? *J. of Archaeol Sci* 20:689-704.
- Bocquet JP (1977) Perspectives paléodémographiques. Ph.D. Thesis, Ecole des Hautes Etudes en Sciences Sociales, Paris.
- Bocquet JP (1978) Estimation methods of age at death in adult skeleton and demographic studies of the population of the past. In MD Garralda and RM Grande (ed.): *I Simposio de Antropologia Biologica de España*. Madrid: 37-47.
- Bocquet JP and Masset C (1977) Estimateurs en Paléodémographie. *L'Homme XVII* 4:65-90.
- Bocquet JP, Maia Neto MA, Tavares da Rocha MA and Xavier de Morais MH (1978) Estimation de l'âge au décès des squelettes d'adultes par régressions multiples. *Contrib. Estud. Antropol. Portuguesa, X* 3:107-167.
- Bocquet-Appel JP and Xavier de Morais MH (1987) *Anthropologie et histoire: Un essai de reconstitution de la variation biologique de la population portugaise au XIX<sup>em</sup> siècle*. Paris: Fontes Documentais Portugueses XXII, Fondation Calouste Gulbenkian.
- Bocquet-Appel JP (1986) Once upon a time: paleodemog-

- raphy. Mitteil. Berlin Gessell. Anthropol. Ethnol. Urges 7:127-133.
- Bocquet-Appel JP (1994) Estimating the average for an unknown age distribution in Anthropology. S Borgognini-Tarli, M Di Bacco, and E Pacciani (eds.): *Statistical Tools in Human Biology*. Singapore: World Scientific pp. 197-202.
- Bocquet-Appel JP and Masset C (1982) Farewell to paleodemography. *J. Hum. Evol.* 11:321-333.
- Bocquet-Appel JP and Masset C (1985) Paleodemography: Resurrection or ghost? *J. Hum. Evol.* 14:107-111.
- Bourgeois-Pichat J (1994) La dynamique des Populations. Populations stables, semi-stables, quasi-stables. Paris: INED, Presses universitaires de France.
- Buikstra JE and Konigsberg LW (1985) Paleodemography: Critiques and controversies. *Am. Anthropol.* 87:316-333.
- Coale AJ and Demeny P (1983) Regional model life tables and stable population. New York: Office of Population Research, Academic Press.
- Cutler RC (1975) Evolution of human longevity and the genetic complexity governing aging rate. *Proc. Natl. Acad. of Sci. U S A* 72/11:4664-4668.
- Deming WE and Stephan FF (1940) On a least squares adjustment of a sampled frequency table when the expected marginal totals are known. *Ann. Math. Statist.* 11:427-444.
- Fienberg SE (1970) An iterative procedure for estimation in contingency tables. *Ann. Math. Statist.* 41,3:907-917.
- Gage TB (1990) Variation and classification of human age patterns of mortality: analysis using competing hazards models. *Hum. Biol.* 62:589-617.
- Gage TB (1991) Causes of death and the components of mortality: Testing the biological interpretations of a competing hazards model. *Am. J. Hum. Biol.* 3:289-300.
- Guy H and Wabont M (1994) Contrôle des Estimateurs paléodémographiques au moyen de registres paroissiaux d'Ancien Régime. *Bull. Mem. Soc. Anthropol. Paris.*
- Haberman SJ (1974) The analysis of frequency data. Chicago: The University of Chicago Press.
- Jackes MK (1985) Pubic symphysis age distributions. *Am. J. Phys. Anthropol.* 68:281-299.
- Jackes M (1992) Paleodemography: Problems and techniques. In S.R. Saunders and M.A. Katzenberg (eds.): *Skeletal Biology of Past Peoples. Research Methods*. New York: Wiley Liss, pp. 189-224.
- Jackes M (1993) On paradox and osteology. *Curr. Anthropol.*, 34,4:435-438.
- Johansson SR and Horowitz S (1986) Estimating mortality in skeletal populations: influence of the growth rate on the interpretation of levels and trends during the transition to agriculture. *Am. J. Phys. Anthropol.* 71:233-250.
- Kennedy JW and Gentle JE (1980) Statistical computing. New York: Marcel Dekker.
- Kimura DK and Chikuni S (1987) Mixtures of empirical distributions: An iterative application of the age length key. *Biometrics* 43:23-35.
- Konigsberg LW and Frankenberg SR (1992) Estimation of Age structure in Anthropological Demography. *Am. J. Phys. Anthropol.* 89:235-256.
- Langenscheidt F (1985) Methodenkritische Untersuchungen zur Paläodemographie am Beispiel zweier fränkischer Gräberfelder. Wiesbaden, Germany: Wiesbaden Bundesinstitut für Bevölkerungsforschung.
- Ledermann S (1969) Nouvelles tables-types de mortalité. Paris: I.N.E.D., Travaux et Documents, 53, Presses Universitaires de France.
- Masset C (1973) La démographie des populations inhumées. *L'Homme* 13,4:95-131.
- Masset C (1982) Estimation de l'âge par les sutures crâniennes. Sc. D., Université de Paris VII.
- Masset C (1989) Age estimation on the basis of cranial sutures. In MY Iscan (eds.): *Age Markers in the Human Skeleton*. Springfield, CC Thomas, br 71-103.
- Masset C and Parzys B (1985) Démographie des cimetières? Incertitude statistique des estimateurs en paléodémographie. *L'Homme*, XXV,2:147-154.
- Meyer MM (1982): Transforming contingency tables. *Ann. of statis.*: 10,4: 1172-1181.
- Molleson TI, Cohen P (1990) The progression of dental attrition stages used for age assessment. *J. of Archaeological Science*, 17:363-371.
- Molleson T, Cox M (1993): Spitafields project. York: Council for British Archaeology.
- Piontek J and Weber A (1990) Controversy on paleodemography. *Int. J. Anthropol.* 5,1:71-83.
- Sacher GA (1975) Maturation and longevity in relation to cranial capacity. In RH Tuttle (ed.): *Primate Functional Morphology and Evolution*. The Hague: Mouton, pp. 417-441.
- Siler W (1979) A competing-risk model for animal mortality. *Ecology* 60:750-757.
- Siler W (1983) Parameters of mortality in human populations with widely varying life spans. *Stat. Med.* 2:373-380.
- Van Gerven DP and Armelagos GJ (1983) "Farewell to paleodemography?", Rumors of its death have been greatly exaggerated. *J. Hum. Evol.* 12:353-360.
- Wood JW, Milner GR, Harpending HC, and Weiss KM (1992) The Osteological Paradox. *Curr. Anthropol.* 33,4:343-370.